

A Machine-learning Approach to Integral Field Unit Spectroscopy Observations. II. H ii Region Line Ratios

Carter Rhea, Laurie Rousseau-Nepton, Simon Prunet, Myriam Prasow-Émond, Julie Hlavacek-Larrondo, Natalia Vale Asari, Kathryn Grasha, Laurence Perreault-Levasseur, Julie Hlavacek-Larr

▶ **To cite this version:**

Carter Rhea, Laurie Rousseau-Nepton, Simon Prunet, Myriam Prasow-Émond, Julie Hlavacek-Larrondo, et al.. A Machine-learning Approach to Integral Field Unit Spectroscopy Observations. II. H ii Region Line Ratios. The Astrophysical Journal, 2021, 910 (2), pp.129. 10.3847/1538-4357/abe627 . hal-03376922

HAL Id: hal-03376922
https://cnrs.hal.science/hal-03376922

Submitted on 18 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Draft version February 15, 2021
Typeset using LaTeX **twocolumn** style in AASTeX63

arXiv:2102.06230v1 [astro-ph.GA] 11 Feb 2021

A Machine Learning Approach to Integral Field Unit Spectroscopy Observations: II. HII Region Line Ratios

Carter Rhea [1], Laurie Rousseau-Nepton [2], Simon Prunet [2,3], Myriam Prasow-Émond [1], Julie Hlavacek-Larrondo [1], Natalia Vale Asari [4,5,6], Kathryn Grasha [7] and Laurence Perreault-Levasseur [1,8,9]

[1] *Département de Physique, Université de Montréal, Succ. Centre-Ville, Montréal, Québec, H3C 3J7, Canada*
[2] *Canada-France-Hawaii Telescope, Kamuela, HI, United States*
[3] *Laboratoire Lagrange, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Parc Valrose, 06104 Nice Cedex 2, France*
[4] *Departamento de Física–CFM, Universidade Federal de Santa Catarina, C.P. 476, 88040-900, Florianópolis, SC, Brazil*
[5] *School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*
[6] *Royal Society–Newton Advanced Fellowship*
[7] *Research School of Astronomy and Astrophysics, Australian National University, Weston Creek, ACT 2611, Australia*
[8] *Mila - Quebec Artificial Intelligence Institute, Montreal, Québec, Canada*
[9] *Center for Computational Astrophysics, Flatiron Institute, New York, USA*

(Received; Revised; Accepted February 15, 2021)

Submitted to ApJ

ABSTRACT

In the first paper of this series (Rhea et al. 2020), we demonstrated that neural networks can robustly and efficiently estimate kinematic parameters for optical emission-line spectra taken by SITELLE at the Canada-France-Hawaii Telescope. This paper expands upon this notion by developing an artificial neural network to estimate the line ratios of strong emission-lines present in the SN1, SN2, and SN3 filters of SITELLE. We construct a set of 50,000 synthetic spectra using line ratios taken from the Mexican Million Model database replicating H II regions. Residual analysis of the network on the test set reveals the network's ability to apply tight constraints to the line ratios. We verified the network's efficacy by constructing an activation map, checking the [N II] doublet fixed ratio, and applying a standard k-fold cross-correlation. Additionally, we apply the network to SITELLE observations of M33; the residuals between the algorithm's estimates and values calculated using standard fitting methods show general agreement. Moreover, the neural network reduces the computational costs by two orders of magnitude. Although standard fitting routines do consistently well depending on the signal-to-noise ratio of the spectral features, the neural network can also excels at predictions in the low signal-to-noise regime within the controlled environment of the training set as well as on observed data when the source spectral properties are well constrained by models. These results reinforce the power of machine learning in spectral analysis.

Keywords: Machine Learning; ISM; Galaxies

1. INTRODUCTION

Emission-line nebulae are a critical part of our understanding of galactic evolution and radiative processes; thus, they are a primary targets of observation in extragalactic astronomy (Kennicutt 1984; Veilleux & Osterbrock 1987; Kewley et al. 2019). H II regions form from clumps of gas in the interstellar medium (ISM) when young O/B stars irradiate the surrounding environment (e.g. Franco et al. 2000; Osterbrock & Ferland 1989; Shields 1990). The region becomes either partially or fully ionized depending on the budget and hardness of ionizing photons, the morphology and the total mass of the mother cloud. H II regions are primarily composed of Hydrogen and Helium; however, they

Corresponding author: Carter Rhea
carterrhea@astro.umontreal.ca

contain non-negligible amounts of metals (e.g. Shields & Tinsley 1976; Garnett & Shields 1987; Kennicutt & Oey 1993; Oey & Kennicutt 1993). Through recombination and collisonal processes between the ionized atoms and the free electrons, the nebula emits characteristic strong emission-lines which indicate their underlying chemical structure (e.g. Baldwin et al. 1981; Crawford et al. 1999; Kewley et al. 2001; Kewley et al. 2006; Kewley et al. 2019). In the optical, the primary emission lines include, but are not limited to, the Balmer series (i.e. H $\alpha\lambda6563$, and H $\beta\lambda4861$), ionized oxygen (i.e. [O II]$\lambda\lambda3726, 3729$, and [O III]$\lambda\lambda4959, 5007$), ionized nitrogen (i.e. [N II]$\lambda\lambda6548, 6583$), and ionized sulfur (i.e. [S II]$\lambda\lambda6716, 6731$). By studying the relative ratios of these lines, the primary emission and ionization mechanism can be determined. Sources that power ionized emission differ and the regions themselves are categorized into several distinct classes: classical H II regions (e.g. Viallefond 1985; Melnick et al. 1987), shock induced regions such as supernova remnants (e.g. Fesen et al. 1985; Danziger & Dennefeld 1976), and planetary nebulae (e.g. Miller 1974; Oserbrock 1964). Several diagnostics exist to categorize emission nebulae based off their line ratios (e.g. Baldwin et al. 1981; Kewley et al. 2019; Constantin & Vogeley 2006; D'Agostino et al. 2019). Nevertheless, these different diagnostics have limitations due to the multi-parameters physical nature of these objects which is not exempt of degeneracy. With that in mind, we want to test the hypothesis that fitting multiple strong lines directly using model predictions (e.g. CLOUDY, MAPPINGS; Ferland et al. 2017 Allen et al. 2008) that covers typical physical conditions in the gas could help minimize the errors associated on the line parameters (intensity, broadening and velocity), provide a pathway to classification of the nebulae directly from the fit, and potentially significantly increase the computational efficiency of the procedure as well as providing a new angle to estimate uncertainties.

The recent advent of integral field spectroscopy is expanding our knowledge of emission-line nebulae with their increased spatial and spectral resolution; these instruments are pushing the ability of existing analysis tools due to their resolution. (e.g. Sánchez et al. 2012; Leroy et al. 2016; Bundy et al. 2014; Martins et al. 2010). SITELLE, an Imaging Fourier Transform Spectrograph (IFTS) located at the Canada-France-Hawai'i Telescope, is one such instrument. SITELLE has an unrivaled field of view of $11' \times 11'$ and produces data cubes containing over 4 million pixels. Each pixel contains a spectrum. (e.g. Baril et al. 2016; Drissen et al. 2019). SITELLE has a spectral resolution between 1 and 20,000. It's instrumental line shape is described by a cardinal sinc function. It can be convolved with a Gaussian (Martin et al. 2016) to account for natural broadening of the observed lines. Spectral fits must be done cautiously to accurately model the line shape and capture the effects of sidelobes of the sinc, which can greatly influence the estimated flux of an emission-line (e.g. Martin et al. 2012). In many field of astrophysics, it is necessary to model resolved and unresolved (blended) lines accurately. Emission line parameters are the root information used to infer physical properties of locally-resolved nebulae as well as classifying them, but they are also crucial for proper characterisation of galaxies at different redshift. For the later, the line parameters are essential to fully understand how internal and external processes affect galaxies as one cannot disentangle active galactic nucleus, stellar winds, supernovae, etc, multiple feedback processes that affect the whole galaxy in different ways and at different scales.

In this paper we apply an artificial neural network to SITELLE spectra in order to obtain the strong emission-line ratios and validates its results by comparing with the line ratio derived from a standard line fitting techniques. We demonstrate the capability of the machine learning algorithm in low signal-to-noise regimes. In §2, we describe the synthetic data used to train and test the algorithm. It is important to note that although the method is tested here on well resolved objects, it could as well be used on galaxy at higher redshift and on integrated spectra. Additionally, we dissect the neural network and its hyperparameters. In §3, we describe how the network was trained and how it compares with the standard fitting procedure. In §4, we explore the impact of the signal-to-noise on the algorithm in addition to applying it on a real observation of M33, a well resolved local galaxy. We conclude in §5.

2. METHODOLOGY

In the first article of the series, Rhea et al. (2020), we explored the application of a convolutional neural network to calculate the kinematic parameters from SITELLE spectra. In this paper we expand the use of machine learning to calculate another critical set of physical parameters, the line ratios of strong emission-lines. As in the previous paper, the initial step in any machine learning application is to assemble the appropriate training set.

2.1. *Synthetic Data*

In order to facilitate the training of the network used to estimate strong-line ratios, we rely on carefully constructed synthetic spectral data. Synthetic

data for emission-line sources must account for both the instrumental line shape (ILS) of an instrument for a given spectral resolution and the relative intensities and broadening of the lines available through observations. The network was developed for use in the SIGNALS program, which uses SITELLE to observe in three bands: SN1 (365–385 nm), SN2 (480–520 nm), and SN3 (651–685 nm). The synthetic data, and observed data, contain approximately 841 channels in SN3, 219 channels in SN2, and 171 channels in SN1. Each strong emission-line's ILS (Hα(6563)Å, [O II]λ3726, [O II]λ3729, Hβ(4861)Å, [O III]λ4959, [O III]λ5007, [N II]λ6548, [N II]λ6583, (Hα(6563)Å, [S II]λ6716, and [S II]λ6731, was separately modeled using the routine `orb.fit.create_cm1_lines_model` (Martin et al. 2012); `ORB` is the software kernel written for SITELLE, the imaging Fourier Transform Spectrometer on the Canada-France-Hawaii Telescope (Martin et al. 2016; Martin & Drissen 2017; Baril et al. 2016). `ORBS` and `ORCS` build upon the `ORB` kernal and are used for data reduction and data analysis, respectively.

Following the SIGNALS large program instrumental configuration for the observations, we set the maximal spectral resolution in SN1 and SN2 to R$\sim$1000 and in SN3 to R$\sim$5000 (Rousseau-Nepton et al. 2019) to create the training set. Since the resolution varies as a function of location in a SITELLE spectral cube, we allow for variations within our spectral resolution to R$\sim$200 less than the maximal value (e.g. Martin & Drissen 2017). Previous work reveled that this is a key parameter for an accurate training of the algorithm (Rhea et al. 2020). To fully sample the velocity and broadening space expected in the SIGNALS catalog (and in keeping with our previous study), we randomly select the velocity parameter from a uniform distribution between -200 and 200 km s^{-1} and the broadening parameter between 10 and 50 km s^{-1}. These values represent expected values for typical HII regions (e.g. Epinat et al. 2008); additionally, at R$\sim$5000 SITELLE is unable to resolve the broadening parameter below 10 km s^{-1} (Rousseau-Nepton et al. 2019). The resolution, broadening, and velocity were randomly selected with replacement for each synthetic spectra so that we sample the entire parameter space (e.g., James et al. 2013). Furthermore, the signal-to-noise ratio varied randomly between 5 and 30 with respect to the Halpha emission – meaning that Halpha emission is at least 5 times over the noise. However, this does not ensure that all lines are above the noise. The noise was assigned to each spectral channel individually and was randomly sampled from a normal distribution centered around the chosen signal-to-noise ratio with a sigma of 1.

The last element required to create synthetic spectra is the relative amplitude of the strong emission-lines. For each type of nebulae, we used different databases. Following the methodology described in detail in Rhea et al. (2020), relative line amplitude of the HII regions were sampled from the Mexican Million Models database (3MdB; Morisset et al. 2015) BOND simulations (Vale Asari et al. 2016). Following standard procedure, we use 70% of the synthetic data for the training set, 20% for the validation set, and 10% for the test set (e.g., Breiman 2001).

In contrast with our previous study, Rhea et al. 2020, we require all strong lines sampled in the SIGNALS observations to be present. We also add the restriction that all strong emission-lines must have an amplitude equal to or greater than 12% of the Hα amplitude. This threshold corresponds to a signal-to-noise ratio of 3 for the faintest targets in the SIGNALS sample whose Hα surface brightness is approximately 8×10^{-17} erg s^{-1} cm^{-2} arcsec^{-2} (Rousseau-Nepton et al. 2019). While the impact of this threshold is discussed more in-depth in § 4.2, note that consistently, the training, validation, and test sets for the primary CNN used in this paper all adhere to this signal-to-noise constraint. Moreover, we select a Balmer decrement by randomly sampling a value between 2 and 6 from a uniform distribution with a sampling increment of 0.01. We then apply reddening at each wavelength sampled in our spectra following the procedure outlined in § 2.5.

We create 50,000 mock spectra in the form of FITS files which contain the emission-line information (e.g. velocity, broadening, resolution, emission-line fluxes) in addition to the spectrum itself.

2.2. *Artificial Neural Networks*

In this paper, we study the application of an artificial neural network to the problem of strong emission-line ratio estimation. Using the synthetic spectra described in §2.1, we train a network to approximate the function which maps SITELLE spectra to their corresponding line ratios.

2.2.1. *The Algorithm*

Feed Forward artificial neural networks (ANN) contain three principal layers: the input layer, the hidden layer(s), and the output layer (e.g. Hansen & Salamon 1990). The input layer consists of the preproccessed data that will be used to train the network and eventually be fed unseen inputs to make predictions. In this case, the input layer is the combined SN1, SN2, and SN3 SITELLE observations described previously. The output layer consists of line ratio estimates. The hidden layers contain an ensemble of nodes which are parametrized

[S II]λ6731/[S II]λ6716	[S II]/Hα	[N II]/Hα	[N II]/[S II]	[O III]/Hβ	[O II]/Hβ	[O II]/[O III]	Hα/Hβ*
0.73-0.79	0.27-1.18	0.27-2.03	0.14-6.97	0.33-2.47	0.24-6.97	0.60-30.89	1.93-5.99

Table 1: Ranges of line ratio parameters from synthetic spectra. The distribution of the parameters are all skewed heavily to the lower values (with the exception of the [S II] doublet ratio which is evenly distributed over a small range) which represents the likelihood of finding those parameters in the simulations from which these values are obtained. After applying an `arcsinh` transformation, the variables are normally distributed. * The wide range is due to artificially injected dust redenning following the prescription described in §2.5.

by a linear function that takes the input, $\mathbf{x}$, multiplies it by a weight matrix, $\mathbf{w}$, and adds a bias, $\mathbf{b}$. The node is then activated by a predetermined function. A common activation function, the Rectified Linear Unit (ReLU), is employed in every layer of the network used in this paper except for the final layer (e.g. Chen et al. 1990). The ReLU function, $g(w \times x + b) = \max(0, w \times x + b)$, takes the value of the node unless the value is negative; in that case, it takes the value 0. The result is a non-negative value for each node in a layer where the vector-valued function of each layer, l, is denoted as $\mathbf{h}^l(\mathbf{x},\mathbf{b},\mathbf{w})$. In traditional neural networks, such as the network applied in this work, the layers are *fully-connected*; thus, each node in a layer is connected to all nodes in the previous and subsequent layer.

After calculating $\mathbf{h}^l(\mathbf{x},\mathbf{b},\mathbf{w})$ sequentially for each layer, we have a vector-valued output $\mathbf{f}(\mathbf{x},\mathbf{b},\mathbf{w})$. We can calculate the *loss* between the final predictions, $\mathbf{f}$, and the correct outputs, $\mathbf{y}$. We adopt the Huber loss function which is defined as:

$$k_\delta(f, y : w, b)) = \begin{cases} \frac{1}{2}(y - f(w,b))^2 & |y - f(w,b)| \leq \delta \\ \delta|y - f(w,b)| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases} \quad (1)$$

where δ is a tune-able parameter and initialized as 1. The Huber loss function is often employed since it reduces the effects of outliers on the final cost calculation (e.g. Huber 1964).

We then minimize the loss function through back-propagation in which we alter the weights and biases (Hecht-Nielsen 1989):

$$\arg\min_{w,b} k(f, y; w, b) \quad (2)$$

We apply the Adam implementation of the stochastic gradient descent algorithm in order to minimize the loss function (Lechevallier & Saporta 2010; Kingma & Ba 2017). In this manner, we are able to train the network by updating the weights and biases until the loss function is minimized on the training set.

2.3. *The Network and Hyper-Parameters*

In order to determine the structural parameters of the network, we first constructed four networks; The first network had two hidden layers, the second network three hidden layers, and so on. Since the problem is nonlinear, we did not test a single hidden layer network. We note that the training, validation, and test sets used to determine the structural parameters contain only 10000 synthetic spectra. These spectra were not re-used in the final training, validation, and testing ensemble. Initially, each hidden layer contained 1000 nodes. This number is sufficient to test the optimal number of layers (e.g. Sheela & Deepa 2013). Figure 1 demonstrates that the MAPE plateaus for the training set at four layers. However, since the mean squared error (MSE) increase in both the validation and test sets between three and four layers, we chose three layers as the optimal value. Although the figure is not included, the same patterns hold for the mean absolute error. Next, we applied a grid search on the number of hidden neurons allowing each to vary from 64, 128, 256, 512, and 1024 nodes. The MSE was minimized for all sets when the first layer has 1024 nodes, the second layer has 1024 nodes, and

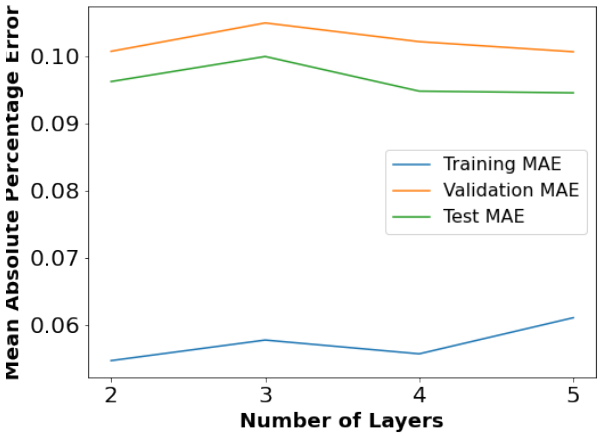


Figure 1: Final mean absolute percentage error after five epochs of training as a function of the number of layers in the neural network. Each layer contains 1000 nodes and is activated by a `relu` function. The graphics show the curve for the mock training, validation, and test sets. We emphasize that the synthetic data used in the structural parameter tuning were not used again.

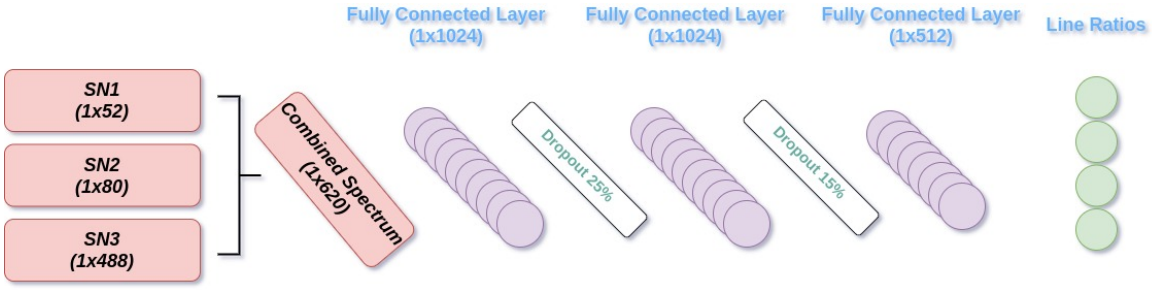


Figure 2: Graphical depiction of the artificial neural network employed in this paper. The spectrum vectors for SN1, SN2, and SN3 are combined to form a single input vector which is then passed through a series of three fully connected layers. A linear activation function takes the final layer and compresses it into a final prediction for each line ratio.

the third layer has 512 nodes. Figure 2 graphically depicts the entire network. The structure of the algorithm is as follows:

1. Concatenate spectral vectors from SN1, SN2, and SN3 as input
2. Fully Connected Layer with ReLU Activation (1x1024)
3. Dropout (25%)
4. Fully Connected Layer with ReLU Activation (1x1024)
5. Dropout (15%)
6. Fully Connected Layer with ReLu Activation (1x512)
7. Fully Connected Output Layer with Linear Activation (1x8)

We determined the optimal hyperparameter values by using the grid search and random search cross-validation techniques implemented in `sklearn`. The batch size's optimal value was determined to be four. Additionally, we adopt a normal random distribution as the initialization of the hidden layer neurons (Thimm & Fiesler 1995; de Castro et al. 1998). The data are normalized using to the maximum value in the the SN3 filter which corresponds to the maximum value in the combined filter. We note that this is not necessarily the amplitude of the Hα line. We use the `tensorflow` implementation of the `ADAM` optimization algorithm.

Initial testing of the algorithm revealed a systematic bias of approximately 10% in the residuals (see §3 for more details). In order to account for the bias, we applied two techniques: a transformation of the target variables and a regularization parameter. Initially, the distributions of the target variables (i.e. the line ratios the network is predicting) were skewed positive; we thus applied several transforms on the target variables (e.g. `log10`, `ln`, `arcsinh`, etc.). Experimentation showed that applying an `arcsinh` transformation resulted in normalized target variable distributions. Furthermore, this reduced the systematic bias considerably (e.g. Zheng & Casari 2008; Kuhn & Johnson 2019). Additionally, we applied a simple `l2` regularization technique with $\lambda = 5 \times 10^{-4}$ (e.g. Phaisangittisagul 2016; van Laarhoven 2017).

In order to increase the accuracy of the method and provide error estimates, we employ a technique known as deep ensembling (Lakshminarayanan et al. 2017). This method leverages the fact that each neural network is independent of the other. We train ten individual networks with the same architecture but with different weight initializations. We then apply each network to the test data individually. We average the classification probabilities and allow the standard deviation to act as an uncertainty estimate.

2.4. *SITELLE Data*

SITELLE observations of the Southwest Field of M33 led by P.I. Laurie Rousseau-Nepton were taken during the Queued Service Observation Period 18B (Program 18BP41). The galaxy was imaged in the three primary filters: SN1, SN2, and SN3. Both the SN1 and SN2 observations were taken at a spectral resolution $R \sim 1000$, while the SN3 observation was taken at a spectral resolution $R \sim 5000$. Although we do not analyse the data from the following two galaxies in detail in this paper, we demonstrate the feasibility of applying the network to two additional SIGNALS galaxies: M95 (Program 19A) and NGC4214 (Program 18A). These observations represent a subsample of the SIGNALS survey. We note that the authors are members of the SIGNALS science team. The raw data were reduced and calibrated using SITELLE's Level-1 reduction software, ORBS (version 3.1.2 Martin et al. 2012).

2.5. *Dereddening*

Since we are using emission-line ratios spanning a large range of wavelengths, we must include the effects of dust attenuation by reddening the spectra (Calzetti et al. 1994; Buat & Xu 1996; Pettini et al. 2001). We calculate dereddening following the standard procedure of postulating an effective dust screen attenuation to obtain the intrinsic emission-line fluxes ($F_{0,\lambda}$),

$$F_{0,\lambda} = F_{\mathrm{obs},\lambda}\, e^{\tau_\lambda} = F_{\mathrm{obs},\lambda}\, e^{\tau_V\, q_\lambda}, \tag{3}$$

where $F_{\mathrm{obs},\lambda}$ is the observed flux, τ_λ is the optical depth at a given wavelength, τ_V is the optical depth in the V-band, and the shape of the dust attenuation curve is parametrized by $q_\lambda \equiv \tau_\lambda/\tau_V$. We adopt the Cardelli et al. (1989) attenuation law with a total-to-selective extinction $R_V = 3.1$. We use the Balmer decrement, defined as $B_d = F_{\mathrm{obs,H}\alpha}/F_{\mathrm{obs,H}\beta}$, to calculate τ_V:

$$\tau_V = \frac{1}{q_{H\beta} - q_{H\alpha}} \ln \frac{B_d}{B_{d,in}}, \tag{4}$$

where $B_{d,in}$ is the intrinsic Balmer decrement. We assume this value to be 2.87, which is appropriate for the Case B, an electronic temperature of 10 000 K and low density (e.g. Osterbrock & Ferland 1989). Dereddening is applied to the line ratios post calculation by the neural network.

3. RESULTS

3.1. *Optimal Synthetic Data*

In this section we discuss the primary results of the paper and address the efficacy of the ANN when applied to optimal SITELLE HII regions sythethic spectra in order to predict the following strong emission-line ratios : [N II]λ6548/[N II]λ6583, [S II]λ6731/[S II]λ6716, ([S II]λ6716+[S II]λ6731)/Hα, [N II]λ6583/Hα, Hα/Hβ, [N II]λ6583/([S II]λ6716+[S II]λ6731), [O III]λ5007/Hβ, ([O II]λ3726+[O II]λ3729)/[O III]λ5007, and ([O II]λ3726+[O II]λ3729)/Hβ.

We train and validate the network using the synthetic data described in §2.1. The mean absolute errors ($\sum_i^N$ `abs`$(y_{true_i} - y_{pred_i})$; N is the number of training and validation input) on the training and validation are 0.0555 and 0.1207, respectively. The algorithm is applied to the test set; Figure 3 shows the relative errors achieved by the network when recovering each emission-line ratio. We note that all line-ratios recovered from SN3 have a low residual standard deviation; this is attributed to the higher spectral resolution in SN3 compared to SN1 and SN2. All error plots reveal approximately Gaussian error distributions with a positive skew. In order to validate these results using the standard fitting techniques, we use the `ORB` routine `fit.fit_lines_in_spectrum`. We individually fit each filter; however, within a given filter, we fit all emission-lines simultaneously and supply the routine with the correct velocity and broadening parameters initially in order to retrieve the best possible fits. The lines were fit with a `sincgauss` function (Martin et al. 2016). Table 2 displays the comparison between the relative errors obtained using standard fitting routines and our ANN on the training set. In these conditions, the network outperforms the standard method for line ratios in all three filters. It is important to note that the relatively high errors in the fits are largely due to signal-to-noise effects (see §4.2 for the discussion). Appendix A contains the residuals binned by signal-to-noise for both the neural network and standard fits. These plots illustrate that the neural network is capable of outperforming a fitting routine in the low SNR (signal-to-noise) regime when compared within the training set environment. Importantly, the Hα/Hβ errors have the potential to be reduced significantly which will lead to higher fidelity dereddening estimates.

3.2. *Noise Classification*

In this section we evaluate the network used to classify a spectrum as noisy or clean; we define a spectrum clean if the SNR of all strong emission-lines is above a certain pre-determined threshold – in this case 5% that of Hα. This value was determined by running our network on the SN3 spectra of varying thresholds (from 1%–20%). The results indicated that below 5%, the ability of the network to recover the line ratios becomes inhibited.

Synthetic spectra for all filters were created in a method identical to that described in § 2.1 except that we removed the requirement that all strong emission-lines have an amplitude equal to at least 12% that of Hα. Spectra were classified as noisy if any emission-line had an amplitude less than 5% that of Hα; otherwise, the spectrum was classified as noiseless. In spectra where a single strong emission-line amplitude was below the chosen threshold, several other lines were also beneath the threshold; thus, this constraint accurately categorizes data into noisy and noiseless. We created 1,000 noisy and 1,000 noiseless spectra with signal-to-noise ratios of Hα varying between 5 and 30.

We use a Decision Tree Classifier, but in order to reduce bias and probabilistic errors, we aggregate several trees into a Random Forest. The data were randomly shuffled; 90% were set aside for training, and 10% for testing. Using 10 estimators (or 10 decision trees), the Random Forest Classifier reports 100% accuracy in classifying the two spectral types resulting in a diagonal confusion matrix. Although reaching perfect accuracy

Fitting Procedure	[S II]λ6731/[S II]λ6716	[S II]/Hα	[N II]/Hα	[N II]/[S II]	[O III]/Hβ	[O II]/Hβ	[O II]/[O III]	Hα/Hβ
`ORB`	426.68%	17.87%	36.67%	163.56%	190.89%	96.30%	558.75%	90.90%
`ORB (SNR>10)`	42.51%	18.74%	19.06%	26.67%	35.82%	55.43%	100.07%	20.14%
`ANN`	0.94%	19.18%	30.242%	31.67%	24.68%	8.99%	15.36%	6.75%

Table 2: Standard deviation calculations for relative errors of strong emission-line ratios. The top row reports values calculated using the standard `ORB` routine while the second row calculates the same values for test set spectra with a signal-to-noise ratio greater than ten; comparatively, the bottom row contains values obtained using our ANN. [S II] refers to the addition of the [S II] doublets: [S II]λ6713 and [S II]λ6731; similarly, [O II] refers to the addition of the two primary [O II] lines: [O II]λ3726 and [O II]λ3729. Note that [O III] refers only to a single line: [O III]λ5007. As expected, the values calculated can be significantly elevated because the test set contains spectra in the extremely low signal-to-noise regime. We discuss the effects of signal-to-noise in detail in §4.2.

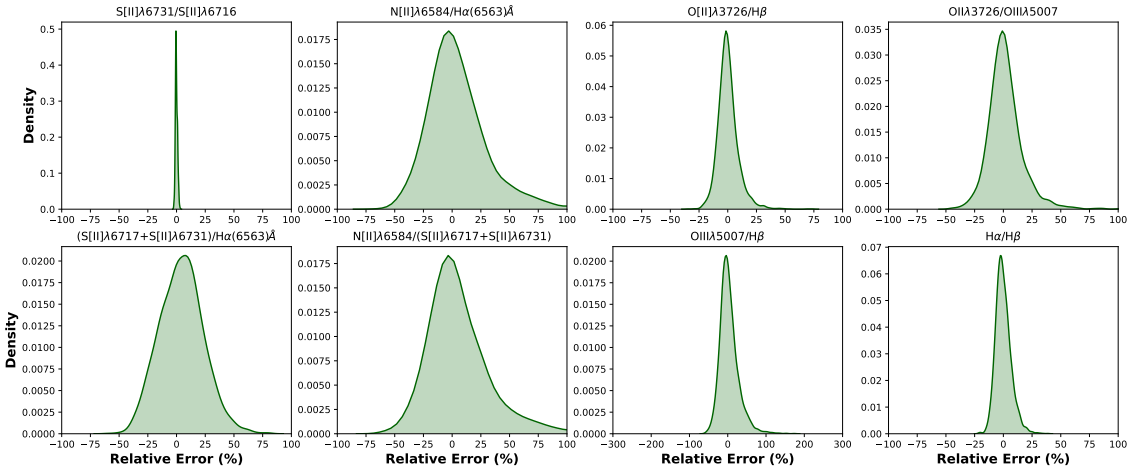


Figure 3: Density plots of line ratio relative errors calculated using the ANN described in §2.2. We note that the errors follow an approximately Gaussian distribution with a non-negligible positive skew.

is relatively rare in machine learning scenarios, the well-defined and simple task allow the Random Forest Classifier to achieve 100% accuracy. Furthermore, if instead 70% of the data is taken for training, the power of the algorithm allows us to simply increase the number of estimators in order to obtain the same perfect accuracy. A future paper will dig deeper into the Noise classification problem.

4. DISCUSSION

4.1. *Verification of the Network*

Although neural networks trained on synthetic data are notoriously difficult to verify, we explore several methods to test whether the network is accurately learning the line ratios and is portable to real data (e.g. Bishop 1994; Krogh & Vedelsby 1995). We apply the following three techniques: k-fold cross-validation, tracking the static [N II]λ6548/[N II]λ6583 ratio, and a saliency map.

We apply a standard k-fold cross-validation by partitioning our synthetic data set into ten equal-sized allotments. We then train and validate the network with nine of the ten partitions allowing the final partition to serve as the test set. This process is repeated until each partition has been served as a test set (e.g. Bengio & Grandvalet 2004; Picard & Cook 1984). This method validates the network's estimates by testing the network for overfitting. Overfitting occurs when the network learns the training set but is unable to generalize to new data. If the network is susceptible to overfitting, then the final accuracy scores of the k-fold cross-validation model will differ significantly. A k-fold cross-validation analysis of our network indicates no overt overfitting in the model since the accuracy are all within several percentage points of one another (Molinaro et al. 2005; Cawley & Talbot 2010).

Additionally, we estimate the line ratio [N II]λ6548/[N II]λ6583 which is expected to be constant

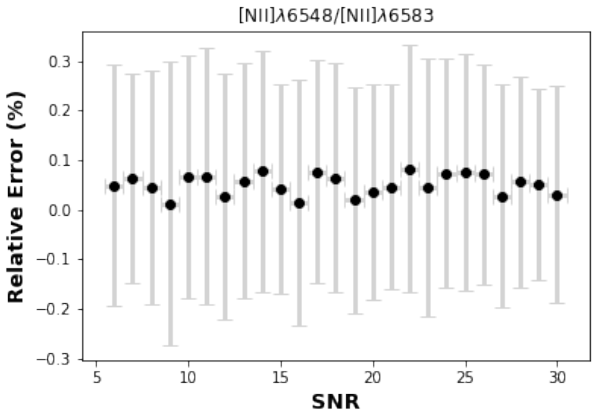


Figure 4: Relative error as a function of signal-to-noise for the [N II] doublet: [N II]λ6548/[N II]λ6583. The black dots represent the median relative error in a given signal-to-noise bin while the grey bars are the 1-σ errors associated with a given bin.

throughout all emission-line nebula and is frequently set to 3 in fitting code (e.g. Schirmer et al. 2013). This is reflected in the synthetic data set. Figure 4 reveals that the calculated [N II] doublet ratio relative error between the true and network-estimated value is between 0 and -1% while the standard deviation is approximately 0.27%. Therefore, the network accurately replicates the static relation between the [N II] lines.

Subsequently, we calculate the saliency map of the network. In order to calculate the saliency map, we calculate the derivative of the input with respect to the output of the neural network: $\frac{\partial[input]}{\partial[output]}$. By multiplying the gradients together, we are left with our desired partial derivative (e.g. Simonyan et al. 2014). We normalize the values to unity in order to compare their relative importance with ease. In this manner, input nodes with a saliency value of zero do not affect the neural network, while input nodes with a saliency value of 1 have the most importance in affecting the neural network's estimation. In Figure 5, we show the saliency values plotted over a reference spectrum with a signal-to-noise of 20; we note that saliency values less than 0.05 are not included in the figure. The saliency map clearly shows that the network prioritizes the amplitude of the emission lines. However, the map also reveals the importance of the regions adjacent to the emission-lines and the areas of the spectrum in between emission-lines. This reflects the importance of the sidelobes of the ILS on the flux ratio estimates.

4.2. *Effects of Varying Signal-to-Noise*

Signal-to-noise constrains the efficacy of traditional fitting methods (e.g. Campbell et al. 1986; Endl & Cochran 2016). In this section, we compare the effects of signal-to-noise on the standard fitting techniques implemented in `ORB` compared to the artificial neural network described in this paper.

In order to study the effects of the signal-to-noise on the efficacy of line ratio estimates, we bin the line ratio residuals as a function of signal-to-noise. A signal-to-noise bin is created at each integer value of the sampled signal-to-noise used to create the synthetic data (5-030). Residuals were calculated by taking the estimated value and subtracted the ground truth and dividing that value by the ground truth value; the value was then multiplied by 100 to make it a percentage. We then removed all outliers; outliers are defined as residual values more than 3-σ off the median value. We then calculated the median value of the remaining set of residuals in each signal-to-noise bin. Errors were calculated as the 1-σ deviations from the median. The plots are shown in appendix A. Figure 7 demonstrates that the residuals do not change as a function of the signal-to-noise when calculated by the neural network. Conversely, the residuals and their associated errors are greatly reduced in high signal-to-noise regimes ($R > 20$) when calculated using standard fitting techniques.

4.3. *Application to M33*

Having demonstrated the feasibility of using a neural network to estimate strong emission-line ratios, we apply our methodology to the Southwest field of M33 studied in our previous article (Rhea et al. 2020). This field contains several previously identified emission region types (classic H II regions, supernova remnants, and planetary nebulae; e.g. Zaritsky et al. 1989; Hodge et al. 1999; Viallefond et al. 1986); additionally, this field is a SIGNALS target. All fits (both from the algorithm developed here and `ORCS`) were run using a computing server located at the CFHT headquarters in Waimea, Hawaii named *iolani*. The server has 2 Intel `XEON E5-2630 v3` CPUs operating at 2.40GHz with 8 cores each. The configuration also has 64 GB of RAM available for computing purposes.

To compare our results with those from the `ORCS` fitting pipeline, we fit each cube seperately using `ORCS`. We use the `fit_lines_in_region` function to fit the strong emission-lines present in a given filter. Within a filter, the lines are fit simultaneously with a single sinc function convolved with a Gaussian which returns each line's flux, velocity, and broadening (Martin & Drissen 2017). `ORCS` fits the observed data to a sinc function convolved with a Gaussian using the Levenberg-Marquardt least squares optimization algorithm. Velocity and broadening priors were determined by fitting a binned (8×8)

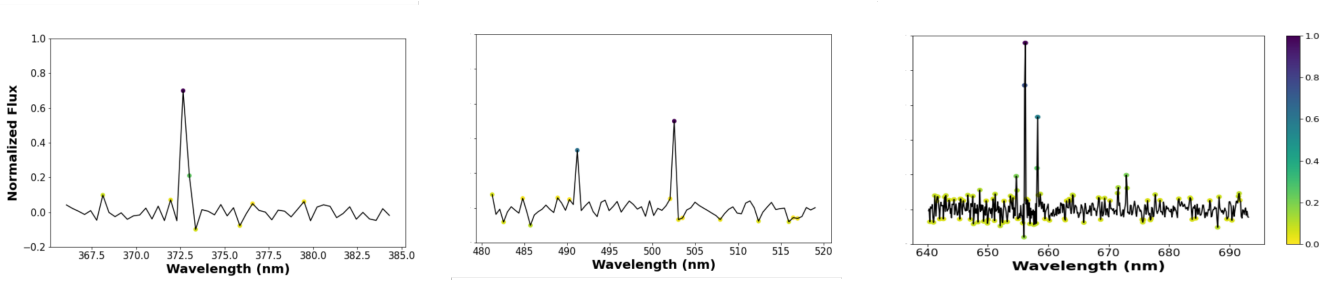


Figure 5: Activation Map of the SN1, SN2, and SN3 filters using the reference spectrum. The relative weights are centered on emission line peaks and surrounding regions reflecting the importance of amplitude and side-lobes on flux ratio estimations. The signal-to-noise of this sample spectrum is approximately 20.

data cube. The final unbinned fits ($\sim$ 4 million spaxels) of SN3 took approximately 11 days, SN2 took 5.5 days, and SN1 took 1.8 day. We note that these calculations are done in parallel. Additionally, we applied the trained neural network to the cube. By comparison, the machine learning methodology requires approximately 18 computing hours, non-parallelized, to complete the entire cube (all three filters). This represents a reduction in computation time by approximately 2 orders of magnitude. Additionally, we applied the network outlined in Rhea et al. (2020) to calculate the velocity and broadening in order to have a more robust timing analysis. We report that the addition of this network did not add any significant amount of time to the overall fitting of the data cube. These results reinforce the notion introduced in Rhea et al. (2020) that by taking a machine learning approach to the analysis of IFU data cubes we can greatly reduce the time required for computation while retaining the accuracy of the results with a proper training.

Although there does not exist a known ground-truth value for these line ratios, we calculate the residuals between the machine learning predictions and standard `ORCS` fits in order to quantify the accuracy of the algorithm. Normalized residual plots for each line ratio can be found in appendix B. A visual inspection of the residual plots reveal that the machine learning algorithm returns similar values to those calculated by `ORCS`. Results deviate most strongly in regions which we show in Rhea et al. (2020) to be best described by multiple emission profiles, regions identified as non-H II regions, and those with a low signal-to-noise ratio. We note, however, that the normalized residual plots (Figures 9 and 10) show general agreement between the standard `ORCS` fits and those calculated by the neural network. Furthermore, discrepancies between the `ORCS` and neural network estimates illustrate the limitations of such an approach when used as a replacement to global fitting algorithms such as `ORCS`. These results further indicate the importance of taking multiple emission profiles into account when modeling – this is the topic of the following paper in the series.

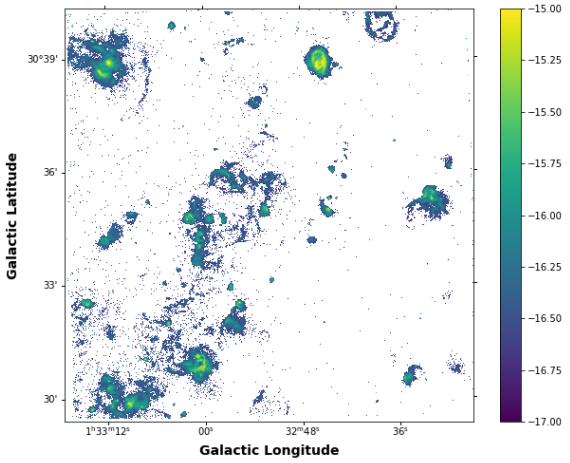


Figure 6: Coadded Hα and [N II]λ6583 emission map. The image illustrates the density of emission-line regions in the outskirts of M33. All pixels with an Hα flux less than 2×10^{-17}erg s^{-1} cm^{-2} are masked out which corresponds to a signal-to-noise ratio of approximately 5.

5. CONCLUSIONS

Applications of machine learning in astronomy are broad: from the estimation of stellar spectral parameters (e.g., Fabbro et al. 2018) to the discovery of objects of interest in extensive astronomical surveys (e.g., Škoda et al. 2020). In this work, we apply an artificial neural network to combined-filter (SN1, SN2, and SN3) SITELLE data representing typical SIGNALS large program observations. The network is designed to calculate important emission-line ratios for H II-like regions which are present in the primary SITELLE filters. We train, validate, and test the algorithm using synthetic data created with the `ORBS` software package. We adopt

physically-derived line amplitudes from the Million Mexican Model Database (Morisset et al. 2015). Our results indicate that the network can potentially constrain the line ratios with greater precision than the standard line fitting technique implemented in `ORCS` if the source spectral properties are well represented in the training set. To demonstrate the applicability of the method beyond synthetic data, we apply the network to the Southwest field of M33. Timing analysis indicates that the network can analyze the entire cube approximately 100 times faster than the standard methods.

These results not only have an impact on the computational aspects of line ratio calculations, but they also carry scientific implications. Although our knowledge of galactic dynamics has expanded considerably over the past several decades, spectroscopic conclusions are restricted by the rigor of the fitting schemes employed and precision of the results. In this paper, we have demonstrated that machine learning algorithms can considerably increase the precision of emission line ratios in both low and high signal-to-noise regimes. This has profound implications on the study of these regions since it will allow stricter categorization using methods such as line-ratio diagnostics in conjunction with BPT diagrams (e.g. Baldwin et al. 1981; Kewley et al. 2006; Kewley et al. 2019). These methods require concise measurements in order to accurately categorize the emission-region type and break any model degeneracies.

Following up on the success of our first report, the work presented here represents the second article in a series of articles covering the application of machine learning algorithms to SITELLE data cubes. Our results have been encouraging in mapping out emission line ratios in pixels dominated by H II region emission serves as a proof-of-concept that using machine learning to identify line fluxes is a viable methodology. We note this work is not meant to be a replacement to global line fitting algorithms. Identifying regions containing multiple, blended emission components, as well as multiple sources of emission with spectral features not represented in a training set remains to be explored. Additionally, example code can be found at https://github.com/sitelle-signals/Pamplemousse.

ACKNOWLEDGMENTS

The authors would like to thank the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. The observations at the CFHT were performed with care and respect from the summit of Maunakea which is a significant cultural and historic site. C. R. acknowledges financial support from the physics department of the Université de Montréal, the MITACS summer scholarship program, and the IVADO doctoral excellence scholarship. J. H.-L. acknowledges support from NSERC via the Discovery grant program, as well as the Canada Research Chair program. NVA acknowledges support of the Royal Society and the Newton Fund via the award of a Royal Society–Newton Advanced Fellowship (grant NAF\R1\180403), and of Fundação de Amparo á Pesquisa e Inovação de Santa Catarina (FAPESC) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

APPENDIX

A. SIGNAL-TO-NOISE AND RESIDUALS

In this section we display the signal-to-noise vs residual plots described in §4.2. Figure 7 displays the signal-to-noise vs residual plots for each line ratio calculated using the machine learning algorithm described in §2.2. Figure 8 shows the signal-to-noise vs residual plots for each line ratio calculated using `ORB` as described in §3.

B. LINE RATIO RESIDUAL PLOTS

This section contains the line ratio residual plots (ORCS fits - ANN estimates; Figures 9 and 10). The two methods are in agreement in regions found to be best described by a single emission profile for each strong-line (see Rhea et al. (2020) for details); in regions best described by two emission profiles, the results differ significantly. In the subsequent paper, we will explore machine learning techniques to determine whether or not emission regions are best described by a single or double emission profile.

REFERENCES

Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., & Kewley, L. J. 2008, The Astrophysical Journal Supplement Series, 178, 20, doi: 10.1086/589652

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, Publications of the Astronomical Society of the Pacific, 93, 5, doi: 10.1086/130766

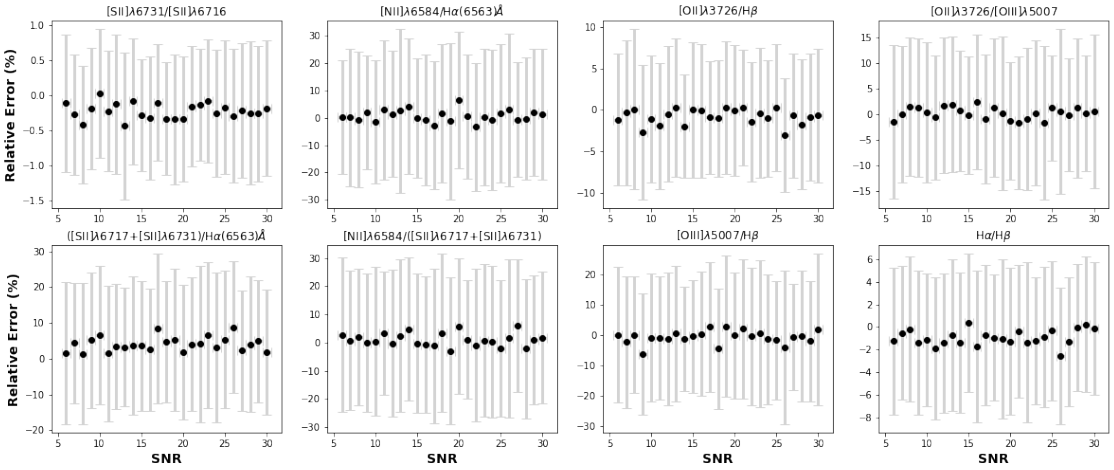


Figure 7: Signal-to-noise ratio vs relative errors for the estimations obtained on the test set using the artificial neural network described in this paper. The signal-to-noise bins were taken at integer intervals. The black dots are the mean residuals. The grey bars represent the 1-sigma errors in a given signal-to-noise bin.

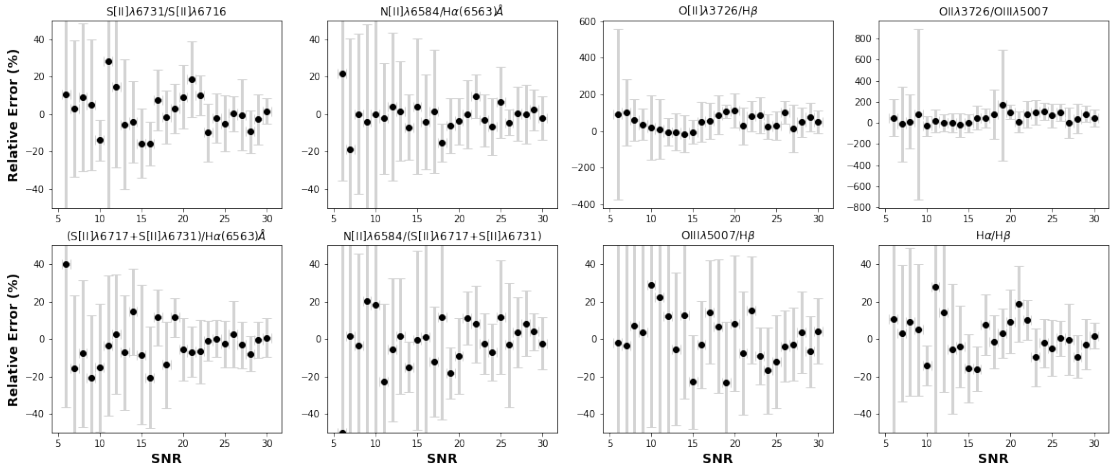


Figure 8: Signal-to-noise ratio vs relative errors for the estimations obtained on the test set using the software package `ORBS`. The black dots are the mean residuals. The signal-to-noise bins were taken at integer intervals. The grey bars represent the 1-sigma errors in a given signal-to-noise bin.

Baril, M., Grandmont, F., Mandar, J., et al. 2016, International Society for Optics and Photonics, 9908

Bengio, Y., & Grandvalet, Y. 2004, Journal of Machine Learning Research, 1089

Bishop, C. M. 1994, IEE Proceedings - Vision, Image and Signal Processing, 141, 217, doi: 10.1049/ip-vis:19941330

Breiman, L. 2001, Machine Learning, 45, 5

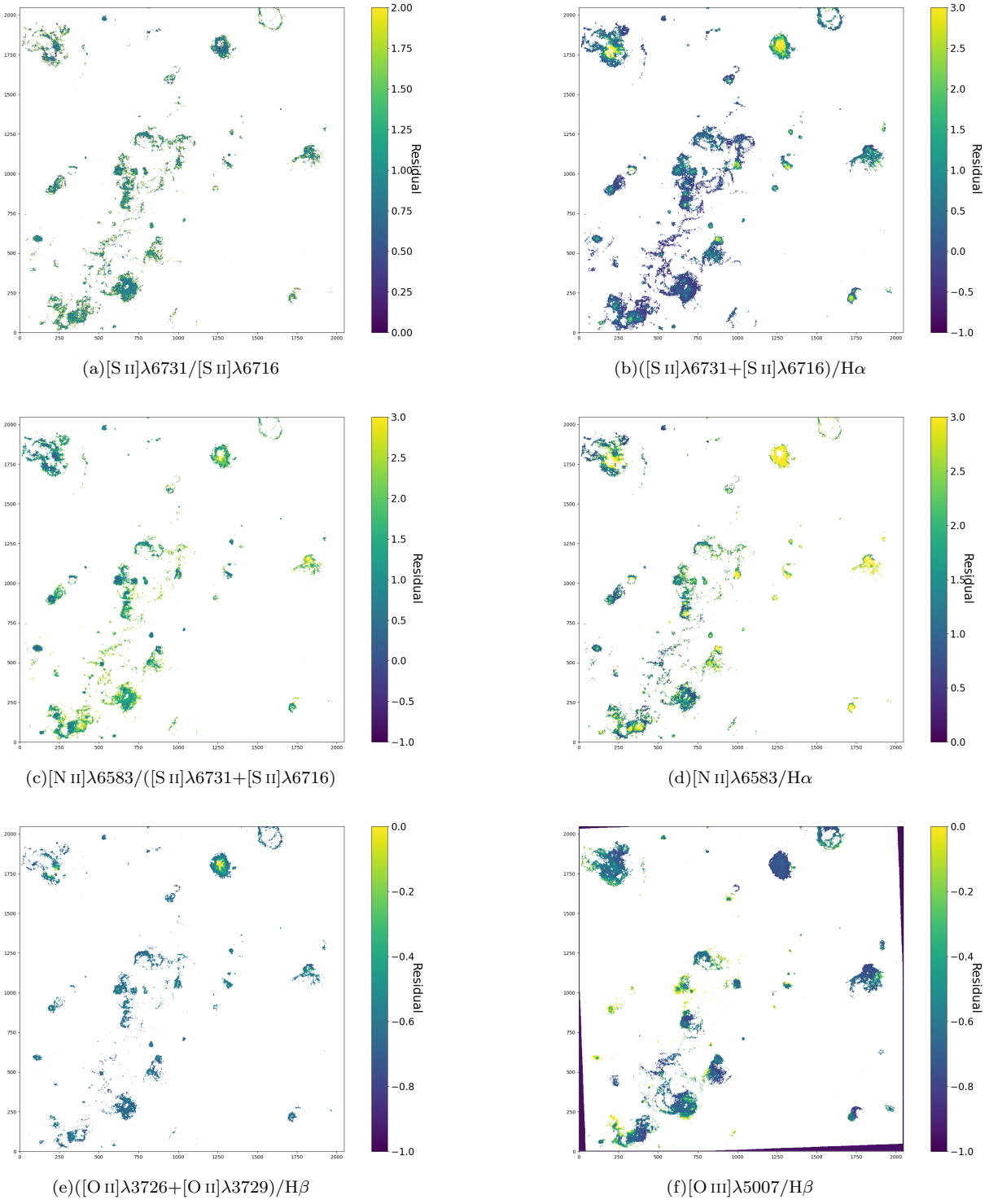
Buat, V., & Xu, C. 1996, Astronomy and Astrophysics, 306, 61. http://adsabs.harvard.edu/abs/1996A%26A...306...61B

Bundy, K., Bershady, M. A., Law, D. R., et al. 2014, The Astrophysical Journal, 798, 7, doi: 10.1088/0004-637X/798/1/7

Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, The Astrophysical Journal, 429, 582, doi: 10.1086/174346

Campbell, A., Terlevich, R., & Melnick, J. 1986, Monthly Notices of the Royal Astronomical Society, 223, 811, doi: 10.1093/mnras/223.4.811

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, The Astrophysical Journal, 345, 245, doi: 10.1086/167900



(a)[S II]λ6731/[S II]λ6716

(b)([S II]λ6731+[S II]λ6716)/Hα

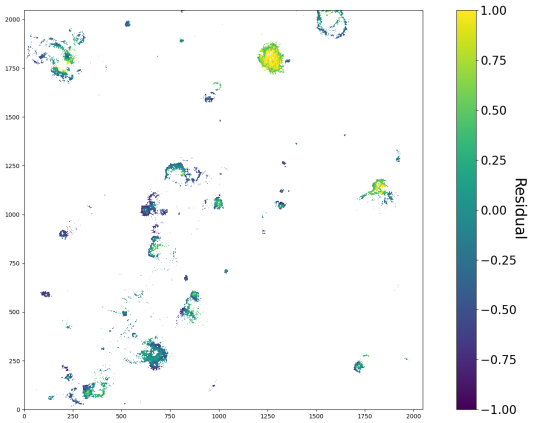
(c)[N II]λ6583/([S II]λ6731+[S II]λ6716)

(d)[N II]λ6583/Hα

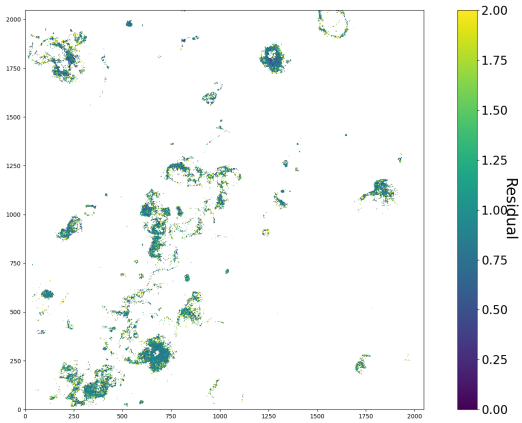
(e)([O II]λ3726+[O II]λ3729)/Hβ

(f)[O III]λ5007/Hβ

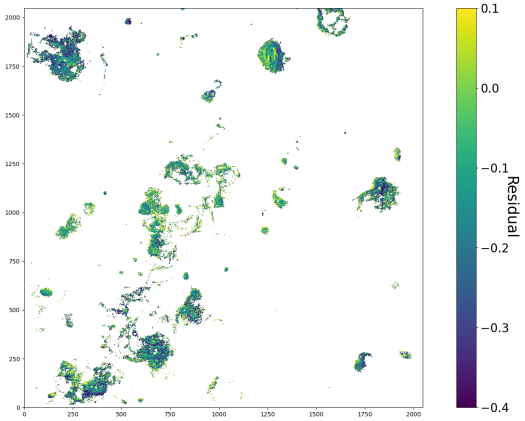
Figure 9: Residual Plots created by taking the difference between the ORCS fits and the values calculated by the artificial neural network for the Southwest Field of M33 normalized by the ORCS fit values. As discussed in the text, regions with large discrepancies between the ORCS and ANN fits are generally either not classic H II regions or are described best by multiple components.



(g)([O II]λ3726+[O II]λ3729)/[O III]λ5007



(h)[S II]λ6731/[S II]λ6716



(i)Hα/Hβ

Figure 10: Extension of Figure 9.

Cawley, G. C., & Talbot, N. L. C. 2010, Journal of Machine Learning Research, 11, 2079

Chen, S., Billings, S., & Grant, P. 1990, Non-linear system identification using neural networks: International Journal of Control: Vol 51, No 6. https://www.tandfonline.com/doi/abs/10.1080/00207179008934126

Constantin, A., & Vogeley, M. S. 2006, The Astrophysical Journal, 650, 727, doi: 10.1086/507087

Crawford, C. S., Allen, S. W., Ebeling, H., Edge, A. C., & Fabian, A. C. 1999, Monthly Notices of the Royal Astronomical Society, 306, 857, doi: 10.1046/j.1365-8711.1999.02583.x

D'Agostino, J. J., Kewley, L. J., Groves, B. A., et al. 2019, Monthly Notices of the Royal Astronomical Society: Letters, 485, L38, doi: 10.1093/mnrasl/slz028

Danziger, I. J., & Dennefeld, M. 1976, Publications of the Astronomical Society of the Pacific, 88, 44, doi: 10.1086/129893

de Castro, L., Iyoda, E., Von Zuben, F., & Gudwin, R. 1998, in Proceedings 5th Brazilian Symposium on Neural Networks (Cat. No.98EX209), 43–48, doi: 10.1109/SBRN.1998.730992

Drissen, L., Martin, T., Rousseau-Nepton, L., et al. 2019, Monthly Notices of the Royal Astronomical Society, 485, 3930, doi: 10.1093/mnras/stz627

Endl, M., & Cochran, W. D. 2016, Publications of the Astronomical Society of the Pacific, 128, 094502, doi: 10.1088/1538-3873/128/967/094502

Epinat, B., Amram, P., & Marcelin, M. 2008, Monthly Notices of the Royal Astronomical Society, 466, doi: 10.1111/j.1365-2966.2008.13796.x

Fabbro, S., Venn, K., O'Briain, T., et al. 2018, Monthly Notices of the Royal Astronomical Society, 475, 2978, doi: 10.1093/mnras/stx3298

Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, arXiv:1705.10877 [astro-ph]. http://arxiv.org/abs/1705.10877

Fesen, R. A., Blair, W. P., & Kirshner, R. P. 1985, The Astrophysical Journal, 292, 29, doi: 10.1086/163130

Franco, J., Kurtz, S. E., García-Segura, G., & Hofner, P. 2000, Astrophysics and Space Science, 272, 169

Garnett, D. R., & Shields, G. A. 1987, Astrophysical Journal, 317, 82

Hansen, L., & Salamon, P. 1990, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, 993, doi: 10.1109/34.58871

Hecht-Nielsen, R. 1989, IJCNN, 13

Hodge, P. W., Balsley, J., Wyder, T. K., & Skelton, B. P. 1999, Publications of the Astronomical Society of the Pacific, 111, 685, doi: 10.1086/316374

Huber, P. J. 1964, Annals of Mathematical Statistics, 35, 73, doi: 10.1214/aoms/1177703732

James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013, An Introduction to Statistical Learning: with Applications in R, 1st edn. (New York: Springer)

Kennicutt, R., & Oey, M. 1993, Revista Mexicana de Astronomía y Astrofísica, 27, 21

Kennicutt, Jr., R. C. 1984, The Astrophysical Journal, 287, 116, doi: 10.1086/162669

Kewley, L., Heislerr, C., & Dopita, M. 2001, Astrophysical Journal Supplement Series, 132, 37

Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, Monthly Notices of the Royal Astronomical Society, 372, 961, doi: 10.1111/j.1365-2966.2006.10859.x

Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, Annual Review of Astronomy and Astrophysics, 57, 511, doi: 10.1146/annurev-astro-081817-051832

Kingma, D. P., & Ba, J. 2017, arXiv:1412.6980 [cs]. http://arxiv.org/abs/1412.6980

Krogh, A., & Vedelsby, J. 1995, in Advances in Neural Information Processing Systems 7, ed. G. Tesauro, D. S. Touretzky, & T. K. Leen (MIT Press), 231–238. http://papers.nips.cc/paper/1001-neural-network-ensembles-cross-validation-and-active-learning.pdf

Kuhn, M., & Johnson, K. 2019, Feature Engineering and Selection: A Practical Approach for Predictive Models (CRC Press)

Lakshminarayanan, B., Pritzel, A., & Blundell, C. 2017, in Advances in Neural Information Processing Systems 30, ed. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Curran Associates, Inc.), 6402–6413. http://papers.nips.cc/paper/7219-simple-and-scalable-predictive-uncertainty-estimation-using-deep

Lechevallier, Y., & Saporta, G., eds. 2010, Proceedings of COMPSTAT'2010 (Heidelberg: Physica-Verlag HD), doi: 10.1007/978-3-7908-2604-3

Leroy, A. K., Hughes, A., Schruba, A., et al. 2016, The Astrophysical Journal, 831, 16, doi: 10.3847/0004-637X/831/1/16

Martin, T., & Drissen, L. 2017, arXiv:1706.03230 [astro-ph]. http://arxiv.org/abs/1706.03230

Martin, T., Drissen, L., & Joncas, G. 2012, in Software and Cyberinfrastructure for Astronomy II, Vol. 8451 (International Society for Optics and Photonics), 84513K, doi: 10.1117/12.925420

Martin, T. B., Prunet, S., & Drissen, L. 2016, Monthly Notices of the Royal Astronomical Society, 463, 4223, doi: 10.1093/mnras/stw2315

Martins, F., Pomarès, M., Deharveng, L., Zavagno, A., & Bouret, J. C. 2010, Astronomy and Astrophysics, 510, A32, doi: 10.1051/0004-6361/200913158

Melnick, J., Terlevich, R., & Moles, M. 1987, Revista Mexicana de Astronomia y Astrofisica, 14, 158

Miller, J. 1974, Annual Review of Astronomy and Astrophysics, 12, 332

Molinaro, A. M., Simon, R., & Pfeiffer, R. M. 2005, Bioinformatics, 21, 3301, doi: 10.1093/bioinformatics/bti499

Morisset, C., Delgado-Inglada, G., & Flores-Fajardo, N. 2015, 19

Oey, M., & Kennicutt, R. 1993, Astrophysical Journal, 411, 137

Oserbrock, D. 1964, Annual Review of Astronomy and Astrophysics, 2, 95, doi: 10.1146/annurev.aa.02.090164.000523

Osterbrock, D., & Ferland, G. 1989, Astrophysics of gaseous nebulae and active galactic nuclei, 1st edn. (Sausalito, CA USA: University Science Books)

Pettini, M., Shapley, A. E., Steidel, C. C., et al. 2001, The Astrophysical Journal, 554, 981, doi: 10.1086/321403

Phaisangittisagul, E. 2016, in 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS) (Bangkok, Thailand: IEEE), 174–179, doi: 10.1109/ISMS.2016.14

Picard, R. R., & Cook, R. D. 1984, Journal of the American Statistical Association, 79, 575, doi: 10.2307/2288403

Rhea, C. L., Rousseau-Nepton, L., Prunet, S., Hlavacek-larrondo, J., & Fabbro, S. 2020, arXiv:2008.08093 [astro-ph]. http://arxiv.org/abs/2008.08093

Rousseau-Nepton, L., Martin, R. P., Robert, C., et al. 2019, Monthly Notices of the Royal Astronomical Society, 489, 5530, doi: 10.1093/mnras/stz2455

Schirmer, M., Diaz, R., Holhjem, K., Levenson, N. A., & Winge, C. 2013, The Astrophysical Journal, 763, 60, doi: 10.1088/0004-637X/763/1/60

Sheela, K. G., & Deepa, S. N. 2013, Review on Methods to Fix Number of Hidden Neurons in Neural Networks, doi: https://doi.org/10.1155/2013/425740

Shields, G. A. 1990, Annual Review of Astronomy and Astrophysics, 28, 525, doi: 10.1146/annurev.aa.28.090190.002521

Shields, G. A., & Tinsley, B. M. 1976, The Astrophysical Journal, 203, 66, doi: 10.1086/154048

Simonyan, K., Vedaldi, A., & Zisserman, A. 2014, Workshop at International Conference on Learning Representations

Sánchez, S. F., Rosales-Ortega, F. F., Marino, R. A., et al. 2012, Astronomy & Astrophysics, 546, A2, doi: 10.1051/0004-6361/201219578

Thimm, G., & Fiesler, E. 1995, in From Natural to Artificial Neural Computation, ed. J. Mira & F. Sandoval, Lecture Notes in Computer Science (Berlin, Heidelberg: Springer), 535–542, doi: 10.1007/3-540-59497-3_220

Vale Asari, N., Stasińska, G., Morisset, C., & Fernandes, R. C. 2016, Monthly Notices of the Royal Astronomical Society, 460, 1739, doi: 10.1093/mnras/stw971

van Laarhoven, T. 2017, arXiv e-prints, 1706, arXiv:1706.05350. http://adsabs.harvard.edu/abs/2017arXiv170605350V

Veilleux, S., & Osterbrock, D. E. 1987, Astrophysical Journal Supplement Series, 63, 295

Viallefond, F. 1985, PhD Dissertation, D. Knuth and TX Thuan, Paris

Viallefond, F., Goss, W. M., van der Hulst, J. M., & Crane, P. C. 1986, Astronomy and Astrophysics Supplement Series, 64, 237. http://adsabs.harvard.edu/abs/1986A%26AS...64..237V

Zaritsky, D., Elston, R., & Hill, J. M. 1989, The Astronomical Journal, 97, 97, doi: 10.1086/114960

Zheng, A., & Casari, A. 2008, Feature Engineering for Machine Learning [Book], 1st edn., Vol. 1 (O'Reilly Media, Inc.). https://www.oreilly.com/library/view/feature-engineering-for/9781491953235/

Škoda, P., Podsztavek, O., & Tvrdík, P. 2020, arXiv e-prints, 2009, arXiv:2009.03219. http://adsabs.harvard.edu/abs/2020arXiv200903219S

-